Original Article

Coding of Central Cancer Registry Industry and Occupation Information: The Texas and Louisiana Experiences

Nancy S. Weiss, MPH, PhD[a]; Sharon P. Cooper, PhD[b]; Christina Socias, DrPH[b]; Ronnie A. Weiss, BA[a]; Vivien W. Chen, MPH, PhD[c]

Abstract: Background: Usual industry and occupation text information have been collected by central cancer registries but few have had the resources to code these data, limiting their usefulness for assessing occupational cancer risks. Study Aims: This project was undertaken to use software available from the National Institute for Occupational Safety and Health (NIOSH) to code industry and occupation information in cancer records reported to the Texas Cancer Registry (TCR) and the Louisiana Tumor Registry (LTR) and to assess the feasibility of its use in ongoing registry operations; to assess the quality of the reported information; and to determine its usefulness in occupational cancer research. Methods: De-identified data files of TCR (n = 103,276) and LTR (n = 26,090) cancer records were obtained for diagnosis years 2010 and 2011, respectively, for cases aged 14 years and older, with industry and occupation text. These data fields were coded to the 2000 US Census Bureau using the NIOSH Industry and Occupation Computerized Coding System (NIOCCS) software at the high level confidence (90% or greater accuracy) and through manual code assignments for records not coded by NIOCCS. Results*:* NIOCCS assigned a code for 37.2% of TCR records and 59.9% of LTR records. Examination of the quality of the coded data found 44.2% of TCR records and 31.1% of LTR records to have missing, unknown, or otherwise insufficient text for assigning a specific industry and occupation code. Additionally, the vague noninformative category of "retired" was reported for 14.9% and 11.2% of TCR and LTR records, respectively. Records with "homemaker/housewife" or those with terms indicating that they never worked represented 7.2% of TCR cases and 9.7% of LTR cases. Excluding the unknown, never worked, and retired categories, no one specific industry or occupation major grouping represented more than 5% of cases in either of the registries. Conclusion*:* NIOCCS is a helpful tool for coding industry and occupation text and continues to improve, but other registry resources are required for implementation into ongoing operations. Improvement in data quality of reported text information in cancer records is paramount to maximize the efficiency of NIOCCS and improve the availability of coded, specific industry and occupation information for occupational cancer research.

Key words: *cancer registry, data quality, industry and occupation, NIOCCS coding software, occupational cancer*

Introduction

Background

According to the National Institute for Occupational Safety and Health (NIOSH) of the Centers for Disease Control and Prevention (CDC), an estimated 40,000 new cancer cases and 20,000 cancer deaths are attributable to occupational exposures in the United States annually.[1] Further, NIOSH estimates that only about 2% of commercially used chemicals have been tested for carcinogenicity.[1] Historically, many human carcinogens have been identified through their association with workplace exposures.[2] There also has been a variety of estimates of the population-attributable fraction of cancer deaths due to occupational exposure, often reported to be in the 2% to 8% range.[3] Although relatively small, the attributable fractions vary considerably by cancer type (eg, higher for lung cancer) and occupation/industry of employment due to potential hazardous exposures (eg, asbestos).[3-4] The burden of occupational cancer also has been shown to disproportionately affect lower-income workers and contribute partly to cancer disparities.[5]

Currently within the United States, 45 states and the District of Columbia receive funding from the CDC for establishing and enhancing population-based cancer registries to document cancer occurrences and associated deaths.[6] The inclusion of industry and occupation information into cancer registries can potentially facilitate occupational cancer prevention and research efforts. Additionally, using systematic coding procedures to standardize available industry and occupation information can improve record precision.[7-8] Researchers have used similar approaches to assess differences in cancer survival by socioeconomic status[9] and to assess occupational cancer risk in California.[10]

Usual industry and occupation information, when available, are required data items for state cancer

[a]Weiss Consulting, Austin, Texas. [b]University of Texas School of Public Health. [c]Louisiana Tumor Registry and Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center.

Address correspondence to Nancy S. Weiss, 5707 Median Road, Austin, TX 78734. Email: nsw1917@att.net.

This project was possible from support by the Centers for Disease Control and Prevention (CDC Cooperative Agreement 5U58DP003915-02), the American Recovery and Reinvestment Act (ARRA) funding through the Office of Secretary Award 200-2008-27957, and the Southwest Center for Occupational and Environmental Health, a NIOSH Education and Research Center, (Grant No. 5T42OH008421). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC.

registries funded under CDC's National Program of Cancer Registries.[11] However, this information has been collected in text form and few cancer registries have had the resources to code these data, limiting their usefulness in assessing occupational cancer risks.

This project was undertaken to assess the feasibility of using software available from NIOSH,[8] newly released in December 2012, to code industry and occupation information in cancer records reported to the Texas Cancer Registry (TCR) and the Louisiana Tumor Registry (LTR). If demonstrated feasible, the NIOSH coding software could potentially be used by cancer registries for enhancing occupational cancer research.

Specific Aims

The aims of this project were:

- to utilize the NIOSH Industry and Occupation Computerized Coding System (NIOCCS) software to code industry and occupation text in TCR and LTR cancer records and assess the feasibility of use in ongoing registry operations
- to assess the quality of the reported industry and occupation information in the registries
- to determine the usefulness of coded data for occupational cancer studies
- to make recommendations for data improvement

Methods

Data Sources

Data from Texas and Louisiana statewide cancer registries were used. These 2 registries were selected because they share some similarities but also have diverse registry operations. As part of the Gulf Coast, both Texas and Louisiana have an abundance of petrochemical extraction and refinery industries as well as numerous chemical plants. Texas is one of the largest states in the nation, with a population of 25.2 million and an annual cancer caseload (new cancer cases) of approximately 108,000.[12] Thus, Texas poses challenges in data collection, including industry and occupation information. The TCR first achieved statewide cancer reporting coverage and national certification in 2006. It receives both state and CDC funding. In contrast, Louisiana is a medium-sized state with a population of about 4.2 million and an annual cancer caseload of approximately 25,000.[13] The LTR is a well-established registry with statewide coverage achieved in 1988. It was nationally certified in 1998, and in 2001 became one of the National Cancer Institute (NCI) Surveillance, Epidemiology and End Results (SEER) registries.[14] Although the LTR receives strong state support and funding from both CDC and the NCI, as for Texas, no ongoing funds have been designated for coding industry and occupation information. This coding project used special one-time funding to assess the feasibility of utilizing NIOCCS for coding industry and occupation in TCR and LTR cancer records.

De-identified data files were obtained from TCR (n= 103,276) and LTR (n= 26,090) reportable cancer records (including in situ carcinomas and benign tumors of the brain and other areas of the central nervous system) for diagnosis year 2010 or 2011, respectively, for cases ages 14 and older. These data were obtained for coding from each registry via a secure file transfer protocol website (WebPlus). The data files consisted of tumors with 1 record or more for each patient and contained the following variables:

- A unique case identification code (for linking the coded information back to the respective cancer patient at the registry)
- Age
- County (parish) of residence (for clarifying industry locations/types)
- Vital status and follow-up source (indicating the source of the patient's vital status designation; this was available for LTR records only to help determine the source of the industry/occupation information)
- Usual occupation text
- Usual industry text

The retrieval and coding of records occurred in early 2013 for the TCR and in spring 2014 for the LTR.

Records Coding

NIOCCS was first introduced as a Web-based coding system and was beta tested among relevant clinicians, health center facilities, cancer registries, and state vital statistics staffs to improve the coding functionality. NIOCCS assigns codes to industry and occupation text at 3 levels of accuracy (or confidence): 30%, 60%, and 90%.[8] The lower confidence (30%) yields the highest number of successfully coded records but also returns more errors. By comparison, the highest confidence yields fewer successfully coded records but with fewer errors. When NIOCCS is unable to assign a code, the field is left blank for manual review. NIOCCS also provides a separate single-code search function where text is manually entered into a search field and NIOCCS provides suggested codes.

The industry and occupation fields were coded to the 2000 US Census Bureau industry and occupation codes using the NIOCCS software publicly available through NIOSH.[8] We selected the high level confidence level (90% or greater accuracy) for automatic code assignments, and records coded by NIOCCS were flagged. Those records that remained uncoded (blank) after the NIOCCS high-level confidence submission were then manually assigned codes by NIOSH-trained staff using either the NIOCCS' single record lookup or hard-copy coding manuals.

Internet searches were performed for company names listed in the industry field to gain sufficient information to code accordingly. NIOCCS and manually assigned industry and occupation codes were reviewed for coding accuracy and consistency for all LTR records and a sample of TCR records. Each industry and occupation code was checked for acceptable code ranges. NIOCCS autocoded records were changed if coding errors were identified or if a more specific code could be assigned; these were then flagged as having been manually coded. The version of NIOCCS used for coding the LTR records in 2014 differed from the earlier version released in 2012 that was used for the TCR records; this later version reflected software improvements.

Data Analyses

Excel files of NIOCCS coded records were extracted from the software and then supplemented with manual code assignments. Microsoft Access[15] was used to create separate databases from the coded files. Queries were run to produce the frequency results shown.

Results

Of the 103,276 TCR records, 37.2% were assigned an industry and occupation code by the software. For the LTR, 59.9% of the 26,090 records were coded by NIOCCS (Table 1). A larger proportion of TCR's 2010 records (14.5%) were missing both industry and occupation text compared with LTR's 2011 records (0.8%) (data not shown). When further examining the quality of the coded data, 44.2% of TCR records and 31.1% of LTR records had missing, unknown, or insufficient text to assign a specific industry and occupation code (Table 2). Additionally, the vague, noninformative category of "retired" was reported for 14.9% and 11.2% of TCR and LTR records, respectively. Records indicating "homemaker/housewife" or those with terms indicating that they never worked represented 7.2% of TCR cases and 9.7% of LTR cases.

Tables 3 through 6 present a breakdown of industry and occupation by major groupings for the 2 registries. With the exception of records with the text mentioned above, no 1 specific industry or occupation major grouping represented more than 5% of cases in either of the registries. Excluding the unknown, never worked, and retired categories, health care and social assistance was the leading industry in both Texas and Louisiana, representing 3.7% and 4.9% of cases, respectively (Tables 3 and 5). Office and administrative support and construction and extraction were the leading occupations for Texas and Louisiana, respectively (Tables 4 and 6).

Table 1. Method of Industry and Occupation Code Assignments for Texas and Louisiana Cancer Records

	Texas 2010 (N = 103,276)		*Louisiana 2011 (N = 26,090)*	
	No.	%	No.	%
Records with NIOCCS assigned industry and occupation codes	38,441*	37.2	15,615	59.9
Records with manual assigned industry and occupation codes	64,835	62.8	10,475	40.1

* Includes 14,976 (14.5%) Texas records with blank occupation and industry information not originally coded by earlier version of software but would be assigned codes by upgraded version used for Louisiana records.

Table 2. Data Quality of Reported Industry and Occupation Information in Texas and Louisiana Cancer Records

	Texas 2010 (N = 103,276)		*Louisiana 2011 (N = 26,090)*	
	No.	%	No.	%
Records with missing/ unknown/non-specific industry and occupation text (codes 999, 990)*	45,666	44.2	8,125	31.1
Records with industry and occupation text of retired (codes 988, 906)	15,432	14.9	2,921	11.2
Records with industry and occupation text of homemaker (codes 989, 901)	3,790	3.7	1,645	6.3
Records with industry and occupation text indicating never worked (codes 989, 910)	3,635	3.5	900	3.4
All other records	34,753	33.7	12,499	47.9

* 2000 US Census Bureau codes.

Discussion

The NIOCCS software provides a useful means of coding industry and occupation text information and has shown improvement since its first release in 2012. One such improvement was the ability of NIOCCS to code records with missing (blank) industry and occupation text to "unknown" for LTR records, whereas the earlier version of the software had required manual exclusion of these records prior to coding of the TCR records.

The initial processing of the large volume of TCR records through NIOCCS (prior to coding LTR records) resulted in identification of corrected and new industry and occupation code assignments, which were incorporated into the software. This resulted in improvement of its performance and possibly contributed to the higher percentage of NIOCCS-assigned codes for Louisiana compared with Texas. Greater use of the NIOCCS software and user feedback to its developers can result in further refinement of its coding engine and continued improvement of the software.

Overall, the benefits of using the NIOCCS software are numerous, including the availability of the software free of charge with NIOSH-supported training in industry and occupation coding and use of the software, the Web-based user-friendly format, the capability for an online industry and occupation look-up with a Census Alphabetical Index, and a crosswalk for coding schemes 2000, 2002, and 2010.[8] Although counts are not available for TCR records coding, only 327 LTR records (2.1%) had a NIOCCS high-level assigned industry and/or occupation code that required correction.

The biggest issue impacting the ability of the NIOCCS software to code industry and occupation was the poor quality of the text fields in the cancer record. These fields contained misspellings, abbreviations, acronyms, transposition of industry/occupation, or insufficient or vague information. Over 60% of Texas records and 40% of

Table 3. Frequency of Industry, Texas Cancer Registry Records, 2010 (Total Record Count, 103,276)			
Industry	*Codes**	*No.*	*%*
Agriculture, Forestry, Fishing, and Hunting	017-029	775	0.8
Mining	037-049	714	0.7
Utilities	057-069	336	0.3
Construction	077	2,241	2.2
Manufacturing	107-399	3,143	3.0
Wholesale Trade	407-459	836	0.8
Retail Trade	467-579	2,889	2.8
Transportation and Warehousing	607-639	2,265	2.2
Information	647-679	617	0.6
Finance and Insurance	687-699	1,466	1.4
Real Estate and Rental and Leasing	707-719	608	0.6
Professional, Scientific, and Technical Services	727-749	2,133	2.1
Management of Companies and Enterprises	757	119	0.1
Administrative and Support & Waste Management Services	758-779	1,424	1.4
Educational Services	786-789	3,722	3.6
Health Care and Social Assistance	797-847	3,817	3.7
Arts, Entertainment, and Recreation	856-859	348	0.3
Accommodation and Food Services	866-869	1,044	1.0
Other Services (Except Public Administration)	877-929	1,904	1.8
Public Administration and Active Duty Military	937-987	2,788	2.7
Retired	988	15,460	15.0
Never Worked, Disabled, Student, Homemaker, Housewife	989	7,832	7.6
Blank, Unknown, Don't Know, NA, Non-specific	999	46,795	45.3

* 2000 US Census Bureau codes.

Louisiana records required labor-intensive manual coding. For this project, the number of records manually coded by an experienced coder averaged 40 to 50 records per hour. NIOCCS record processing times for coding data files also improved greatly from an average of 86 records per minute for TCR records to 256 records per minute for LTR records. As NIOCCS coding and the quality of records improves, the burden of manual coding will drastically decrease.

We found that the quality of industry and occupation information reported in cancer records was better for Louisiana than for Texas, including lower percentages of missing information and records designated only as "retired." This may be the result of the LTR's being a more established statewide registry (and a SEER registry) where the completeness and quality of industry and occupation have been emphasized for recent years. Texas is a larger and newer registry with other registry operations and quality control issues of higher priority than industry and occupation. There also might be facility-related differences in the availability of employment information in the medical record, or differences in the abstractors' ability to identify and record this information. Results of this study can be used by both states for improving the collection of these data fields.

It is not known whether the large percentages of records coded as "unknown" or "retired" for the TCR (59.1%) and the LTR (42.3%) are the result of missing/nonspecific information in the source medical record, or failure by the cancer abstractor to document this information. A study of the quality of industry and occupation data in New Hampshire central cancer registry records reported a significant improvement in the availability of specific industry and occupation information after review of respective medical records; the proportion of reports with no industry or occupation data declined from 77% to 12%.[16] Additionally, this study found improvement in the proportion of unknown industry and occupation reported after training of cancer registrars.[16]

Cancer registry records are linked annually with death records to identify patients who have died for the purposes of updating information on causes of death and missing demographic variables, and identifying unreported cancer cases. TCR and LTR cancer records with either missing or unknown industry and occupation fields were updated with better information when available from the death certificate record. To assess if the source of industry and occupation information impacted the quality of the text data, we examined the percentage of unknown/nonspecified industry and

Table 4. Frequency of Occupation, Texas Cancer Registry Records, 2010 (Total Record Count, 103,276)			
Occupation	*Codes**	*No.*	*%*
Management	001-043	3,185	3.1
Business and Financial Operations	050-095	1,283	1.2
Computer and Mathematical	100-124	412	0.4
Architecture and Engineering	130-156	787	0.8
Life, Physical, and Social Services	160-196	250	0.2
Community and Social Services	200-206	485	0.5
Legal	210-215	358	0.3
Education, Training, and Library	220-255	2,021	2.0
Arts, Design, Entertainment, Sports & Media	260-296	438	0.4
Healthcare Practitioners and Technical	300-354	1,734	1.7
Healthcare Support	360-365	419	0.4
Protective Service	370-395	724	0.7
Food Preparation and Service Related	400-416	645	0.6
Building, Grounds Cleaning and Maintenance	420-425	989	1.0
Personal Care and Service	430-465	703	0.7
Sales and Related	470-496	2,522	2.4
Office and Administrative Support	500-593	3,251	3.1
Farming, Forestry, and Fishing	600-613	199	0.2
Construction and Extraction	620-694	1,897	1.8
Installation, Repair, and Maintenance	700-762	947	0.9
Production	770-896	1,517	1.5
Transportation and Material Moving	900, 903-904, 911-975	2,013	1.9
Housewife, Homemaker	901	3,792	3.7
Volunteer	902	9	0.0
Student	905	398	0.4
Retired	906	16,064	15.6
Unable to Work, Child, Disabled	910	3,652	3.5
Military Occupations	980-985	538	0.5
Blank, Unknown, Don't Know, NA, Nonspecific	990	52,044	50.4

* 2000 US Census Bureau codes.

occupation (codes 999, 990) by vital status for LTR cases. Of the 8,381 LTR patients known to be deceased, 7.1% of cases compared with 42.5% of patients considered alive (17,709) had an unknown/nonspecified industry and occupation. A similar analysis was conducted for "retired" cases (codes 988, 906) and 3.1% of deceased cases were classified as "retired" compared with 15.0% of living cases. Industry and occupation information from the death certificate significantly improved the completeness and quality of registry data.

Based on our coding study, the amount of missing/insufficient data and poor quality of text currently limit the use of these cancer data for occupational cancer research. However, some of these data could be used to explore patterns in cancer incidence and generate hypotheses for future analytic occupational studies. The coding of these data has identified areas of reporting in need of improvement which should be addressed before implementing NIOCCS coding into ongoing registry operations.

Additionally, depending on the detail of industry and occupation text provided in the cancer record, there may be variability in code assignment by different coders. Unlike the reporting and coding of specific topography and morphology information in the cancer record, most industry and occupation text, when not missing or designated as unknown, is vague, which increases intercoder variability. The accuracy of the information also may differ, but cannot be readily assessed without a review of work history sources. For example, it is hard to distinguish if a cancer record with an industry of "oil and gas" and occupation as "operator" should be coded as oil extraction (code 037), oil refining (code 207), or oil wholesale (code

Table 5. Frequency of Industry, Louisiana Tumor Registry Records, 2011 (Total Record Count, 26,090)			
Industry	*Codes**	*No.*	*%*
Agriculture, Forestry, Fishing, and Hunting	017-029	304	1.2
Mining	037-049	474	1.8
Utilities	057-069	142	0.5
Construction	077	1,243	4.8
Manufacturing	107-399	1,195	4.6
Wholesale Trade	407-459	244	0.9
Retail Trade	467-579	1,001	3.8
Transportation and Warehousing	607-639	802	3.1
Information	647-679	225	0.9
Finance and Insurance	687-699	372	1.4
Real Estate and Rental and Leasing	707-719	172	0.7
Professional, Scientific, and Technical Services	727-749	539	2.1
Management of Companies and Enterprises	757	42	0.2
Administrative and Support & Waste Management Services	758-779	418	1.6
Educational Services	786-789	1,226	4.7
Health Care and Social Assistance	797-847	1,269	4.9
Arts, Entertainment, and Recreation	856-859	194	0.7
Accommodation and Food Services	866-869	421	1.6
Other Services (Except Public Administration)	877-929	744	2.9
Public Administration and Active Duty Military	937-987	1,168	4.5
Retired	988	2,921	11.2
Never Worked, Disabled, Student, Homemaker, Housewife	989	2,634	10.1
Blank, Unknown, Don't Know, NA, Non-specific	999	8,340	32.0

* 2000 US Census Bureau codes.

449). When analyzing such data, a researcher may need to consider use of multiple industry codes to identify the oil and gas industry because of this lack of specificity. NIOSH has developed a coding manual for cancer registrars to help improve the collection of industry and occupation data.[17]

Further, the Office of the National Coordinator for Health Information Technology is currently considering the inclusion of industry and occupation in the certification process for electronic health records (EHRs).[18] We are hopeful that this current proposal for support of standardized inclusion of industry and occupation fields in EHRs, as well as the enthusiastic commitment of NIOSH for continued improvement of the software, will result in improved coded data and future research opportunities to examine work-related contributions to cancer.

Recommendations

There is increasing rationale and endorsement for incorporating industry and occupation information in EHRs to support enhanced patient diagnosis, treatment, and population health. The landmark Institute of Medicine Report in 2011 summarized the large burden of occupational illness and injury in the US including cancer, outlined the opportunities and challenges of incorporating occupational information into EHRs, and discussed the groundbreaking prospect of attaining and extending meaningful use of EHRs to improve patient and community health.[7] Examples of potential uses of incorporating occupation and industry in medical records include improved diagnosis and treatment of work-related diseases, targeted data to inform employer wellness programs, and improved public health reporting and surveillance of occupational injuries and illnesses.[7]

However, the current status of the completeness and quality of industry and occupation fields in medical records needs to be assessed and improved before the promise of these opportunities can be realized. Many of the coding limitations originate with the industry and occupation information available in the reported cancer records; this limitation cannot be corrected by coding software. We recommend that cancer registries better understand the quality of their existing records by auditing cancer records to see if industry and occupation information is truly missing or unknown in the medical record. We also recommend monitoring the quality of incoming industry and occupation text, and providing ongoing training of reporters on the collection of this information. Such training opportunities on documenting and recording accurate and sufficient text should be explored, especially at the state cancer registrars'

Table 6. Frequency of Occupation, Louisiana Tumor Registry Records, 2011 (Total Record Count, 26,090)			
Occupation	*Codes**	*No.*	*%*
Management	001-043	989	3.8
Business and Financial Operations	050-095	330	1.3
Computer and Mathematical	100-124	78	0.3
Architecture and Engineering	130-156	230	0.9
Life, Physical, and Social Services	160-196	120	0.5
Community and Social Services	200-206	162	0.6
Legal	210-215	143	0.5
Education, Training, and Library	220-255	643	2.5
Arts, Design, Entertainment, Sports & Media	260-296	135	0.5
Healthcare Practitioners and Technical	300-354	557	2.1
Healthcare Support	360-365	173	0.7
Protective Service	370-395	278	1.1
Food Preparation and Service Related	400-416	359	1.4
Building, Grounds Cleaning and Maintenance	420-425	421	1.6
Personal Care and Service	430-465	317	1.2
Sales and Related	470-496	865	3.3
Office and Administrative Support	500-593	978	0.7
Farming, Forestry, and Fishing	600-613	122	0.5
Construction and Extraction	620-694	1,275	4.9
Installation, Repair, and Maintenance	700-762	464	1.8
Production	770-896	726	2.8
Transportation and Material Moving	900, 903-904, 911-975	956	3.7
Housewife, Homemaker	901	1,645	6.3
Student	905	90	0.3
Retired	906	3,165	12.1
Unable to Work, Child, Disabled	910	900	3.4
Military Occupations	980-985	281	1.1
Blank, Unknown, Don't Know, NA, Nonspecific	990	9,688	37.1

* 2000 US Census Bureau codes.

association annual meeting that most hospital registrars attend. If possible, providing feedback to the data intake personnel on the type of information to collect in the industry and occupation fields, and examples of such, could help improve the text available for coding.

The assessment and subsequent improvement of data quality is central to the decision whether to implement industry and occupation coding into ongoing cancer registry operations. In order to use NIOCCS software, there needs to be dedicated resources for designated staff trained in industry and occupation coding, even if only using NIOCCS autocoding functions, to understand industries in the registry's state, and to include county of diagnosis to assist with industry Web searches. Compiling a dictionary of company names with the specific type of industry within a state or county would facilitate future coding. The NIOCCS software has application in many health department programs, so there could be efficiency in sharing coding resources and expertise across programs.

The collective effort of researchers, cancer registrars, and the medical community is necessary to improve the collection and reporting of industry and occupation in medical records. Until that happens, the data do not exist for reliable population-based, occupation-related cancer epidemiology research. Finally, we recommend review and approval by the respective human subjects institutional review board if individual cancer records with industry and occupation text are released for coding.

In conclusion, NIOCCS is a helpful tool for coding industry and occupation text and continues to improve, but other registry resources are required for implementation into ongoing operations. Improvement in data quality of reported text information in cancer records is essential to maximize the efficiency of NIOCCS and to thereby improve the availability of industry and occupation codes for population-based occupational cancer research.

Acknowledgements

Thanks are extended to Dr. Melanie Williams, who served as manager of the Texas Cancer Registry during the project. Thanks also are given to Pam Schumacher and Sue Nowlin at NIOSH for coding consultation and continued receipt of feedback for improvement of the NIOCCS software.

References

1. National Institute for Occupational Safety and Health (NIOSH) Division of Surveillance, Hazard Evaluations and Field Studies. Program Description. Cancer, Reproductive, and Cardiovascular Diseases. Centers for Disease Control and Prevention website. **http://www.cdc.gov/niosh/programs/crcd/.** Updated August 22, 2013.
2. Cohen SM, Arnold LL. Chemical carcinogenesis. *Toxicol Sci*. 2011;120(suppl 1):S76-S92.
3. Rushton L, Hutchings S, Driscoll T. Estimating the burden of occupational disease. In: Venalbes KM, ed. *Current Topics in Occupational Epidemiology.* New York, NY: Oxford University Press; 2013:186-189.
4. Field RW, Withers BL. Occupation and environmental causes of lung cancer. *Clin Chest Med*. 2013;33(4):681-703.
5. American Cancer Society. Cancer Facts & Figures 2012. Atlanta, GA: American Cancer Society, 2012:51. **http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-031941.pdf**.
6. National Program of Cancer Registries (NPCR): About the program. Centers for Disease Control and Prevention website. **http://www.cdc.gov/cancer/npcr/about.htm**. Updated August 12, 2013.
7. Institute of Medicine. *Incorporating Occupational Information in Electronic Health Records*. Washington, DC: The National Academies Press; 2011.
8. NIOSH Industry and Occupation Computerized Coding System. Centers for Disease Control and Prevention website. **http://www.cdc.gov/niosh/topics/coding/overview.html**. Accessed April 21, 2015.
9. White A, Coker AL, Du XL, Eggleston KS, Williams M. Racial/ethnic disparities in survival among men diagnosed with prostate cancer in Texas. *Cancer*. 2011;117(5):1080-1088.
10. Luckhaupt SE, Deapen D, Cress R, Schumacher P, Shen R, Calvert GM. Leukemia among male construction workers in California, 1988–2007. *Leuk Lymphoma*. 2012;53(11):2228-2236.
11. North American Association of Central Cancer Registries. Volume II: Data Standards and Dictionary Version 15. **http://naaccr.org/StandardsandRegistryOperations/VolumeII.aspx**.
12. Texas Cancer Registry: statistical data. Texas Department of State Health Services website. **https://www.dshs.state.tx.us/tcr/data.shtm**. Accessed May 12, 2015.
13. Hsieh M, Andrews PA, Maniscalco LS, et al, eds. Cancer in Louisiana, 2007–2011: Cancer in Louisiana, Volume 29. New Orleans, LA: Louisiana Tumor Registry; 2014. **http://sph.lsuhsc.edu/Websites/lsupublichealth/images/pdf/ltr/Vol29.pdf**.
14. National Cancer Institute. Surveillance, Epidemiology and End Results (SEER) Program. **http://seer.cancer.gov/registries/**.
15. Microsoft Office Access 2010 [computer program]. **https://products.office.com/en-us/access**.
16. Armenti KR, Celaya MO, Cherala S, Riddle B, Schumacher PK, Rees, JR. Improving the quality of industry and occupation data at a central cancer registry. *Am J Ind Med*. 2010;53(10):995-1001.
17. National Institute for Occupational Safety and Health. *A Cancer Registrar's Guide to Collecting Industry and Occupation*. DHHS (NIOSH) Publication No. 2011-173. April 2011.
18. 2015 edition health information technology (Health IT) certification criteria, 2015 edition base electronic health record (EHR) definition, and ONC health IT certification program modifications: a proposed rule by the health and human services department on 03/30/2015 work information—industry/occupation data. Federal Register website. **https://www.federalregister.gov/articles/2015/03/30/2015-06612/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base#h-51**.